ARTHRITIS-ASSOCIATED PROTEIN

FIELD OF THE INVENTION

This invention relates to methods and compositions for early diagnosis, monitoring and treatment of cartilage degenerative conditions, including forms of arthritis. In particular, the invention relates to a protein differentially expressed in human subjects suffering from arthritis, such as osteoarthritis (OA) and rheumatoid arthritis (RA) compared to individuals without arthritis, antibodies that recognize this protein, and methods for diagnosing arthritis.

BACKGROUND OF THE INVENTION

Arthritis involves both cartilage breakdown and new bone formation, and in many cases, leads to the loss of joint function. While there are some insights as to the mediators contributing to cartilage remodeling, there is only limited knowledge of the underlying molecular mechanisms of arthritis. At a cellular level, the articular chondrocyte plays a key role in both cartilage breakdown and new bone growth, and this is likely to be reflected in changes in cellular transcriptional activity. The application of genomic approaches to the understanding of the complexities of arthritis promises to yield significant advances in the diagnosis and treatment of this prevalent disease of joint degeneration.

An important priority for the investigation and clinical application of potential therapies for arthritis is the identification of biochemical markers that could be used to assess various aspects of disease activity (Chevalier, X. (1997) *Rev.Rhum. Engl. Ed.* **64**: 562-577.

Lohmander, L.S. (1997) *Baillieres Clin. Rhematol.***11**: 711-726). Current research in this field is focusing on a number of possible surrogate markers of arthritis that reflect metabolic changes in the joint associated with cartilage destruction and remodeling, including hyaluronate, cartilage oligomeric matrix protein, keratan sulfate, metalloproteinase activity, and various cytokines.

To identify gene products involved in human arthritis, the present inventors have analyzed the differences in mRNA populations in diseased cartilage compared to age-matched healthy cartilage using differential display RT-PCR (Hu, S-I, Carozza, M., Klein, M., Nantermet, P., Luk, D., and Crowl, R.M. (1998) *J. Biol. Chem.* **273**: 34406-34412). The inventors have

discovered a large protein, termed "adlican," the presence of which in cartilage is strongly associated with arthritis, particularly osteoarthritis. An anti-peptide antibody was generated against the carboxy terminal sequence and used for immunoblot analysis to detect adlican protein in synovial fluid samples obtained from patients with arthritis.

5

SUMMARY OF THE INVENTION

In a first aspect, the invention provides an isolated DNA comprising a nucleotide sequence as set forth in SEQ ID NO:1. Also provided are isolated DNA's comprising nucleic acid sequences that hybridizes under high stringency conditions to the isolated DNA as set forth in SEQ ID NO:1. In a preferred embodiment, the isolated DNA takes the form of a vector molecule comprising the DNA as set forth in SEQ ID NO:1.

10

In a second aspect, the invention provides an isolated polypeptide with an amino acid sequence as set forth in SEQ ID NO:2. Such a polypeptide, or fragments thereof, is found in the synovial fluid of sufferers of arthritis to a much greater extent than in the synovial fluid of individuals without arthritis. In accordance with this aspect of the invention there are provided novel polypeptides of human origin as well as biologically, diagnostically or therapeutically useful fragments, variants and derivatives thereof, variants and derivatives of the fragments, and analogs of the foregoing.

A third aspect of the present invention encompasses a method for the diagnosis of arthritis in a human which includes detecting the elevated transcription of messenger RNA transcribed from adlican-encoding DNA in cartilage from a human, where such elevated transcription is diagnostic of the human's affliction with arthritis.

20

Another embodiment of the assay aspect of the invention provides a method for the diagnosis of arthritis in a human which requires measuring the amount of a polypeptide that includes or is adlican or fragments of adlican in synovial fluid from a human, where the presence of an elevated amount of the polypeptide or fragments thereof, relative to the amount of the polypeptide or fragments thereof in non-arthritic synovial fluid, is diagnostic of the human's suffering from arthritis.

25

In accordance with one embodiment of this aspect of the invention there is provided anti-sense polynucleotides that regulate transcription of the human adlican gene.

30

It is another object of the invention to provide a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs of the foregoing. In a preferred embodiment of this aspect of the invention there are provided methods for producing the aforementioned human adlican

5 polypeptides comprising culturing host cells having incorporated therein an expression vector containing an exogenously-derived human adlican-encoding polynucleotide under conditions sufficient for expression of human adlican polypeptides in the host and then recovering the expressed polypeptide.

In accordance with another object the invention there are provided products,

10 compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides for research, biological, clinical and therapeutic purposes, inter alia.

In certain additional preferred embodiments of this aspect of the invention there are provided an antibody or a fragment thereof which specifically binds to a polypeptide that comprises the amino acid sequence set forth in SEQ ID NO:2, i.e., adlican. In certain

15 particularly preferred embodiments in this regard, the antibodies are highly selective for human adlican polypeptides or portions of human adlican polypeptides.

In a further aspect, an antibody or fragment thereof is provided that binds to a fragment or portion of the amino acid sequence set forth in SEQ ID NO:2.

In yet another aspect, the invention provides vertebrate cells which can be propagated in

20 vitro and which are capable upon growth in culture of producing a polypeptide that comprises the amino acid sequence set forth in SEQ ID NO:2 or fragments thereof, where the cells contain transcriptional control DNA sequences, other than human adlican transcriptional control sequences, where the transcriptional control sequences control transcription of DNA encoding a polypeptide with the amino acid sequence according to SEQ ID NO: 2 or fragments thereof.

25 In another aspect, the present invention provides a method for producing human adlican polypeptides which comprises culturing a host cell having incorporated therein an expression vector containing an exogenously-derived human adlican-encoding polynucleotide under conditions sufficient for expression of human adlican polypeptides in the host cell, thereby causing the production of an expressed polypeptide, and recovering.the expressed polypeptide.

In yet another aspect of the present invention there are provided kits comprising the components necessary to detect above-normal expression of human adlican polynucleotides or polypeptides in body tissue samples derived from a patient.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a depiction of the cDNA encoding adlican (SEQ ID NO:1).

Figure 2 is a depiction of the amino acid sequence encoded by SEQ ID NO:1 (SEQ ID NO:2).

Figure 3 is an analysis of the amount of adlican mRNA expressed in the cartilage of human osteoarthritis suffers compared with the amount of adlican mRNA expressed in the cartilage of human non-arthritic control subjects using semi-quantitative reverse transcription-PCR. Semi-quantitative reverse-transcription PCR of actin was used as a control.

Figure 4 is an autoradiogram of a gel showing the production of adlican *in vitro* using a reticulocyte lysate system.

Figure 5 is a Western Blot of a gel on which proteins from HEK 293 cells which were transfected with an adlican expression vector were run. Anti-adlican peptide antibodies were used. Lane 1 is cell extract from cells transfected with an adlican expression vector 72 hours post-infection. Lane 2 is from cells 48 hours post-infection. Lane 3 is from cells transfected with a control vector at 48 hours post-infection.

Figure 6 is a Western Blot of a gel on which synovial fluid samples from humans suffering from trauma, osteoarthritis, and rheumatoid arthritis were tested. Anti-adlican peptide antibodies were used.

DETAILED DESCRIPTION OF THE INVENTION

All patent applications, patents and literature references cited herein are hereby incorporated by reference in their entirety.

In practicing the present invention, many conventional techniques in molecular biology, microbiology, and recombinant DNA are used. These techniques are well known and are explained in, for example, Current Protocols in Molecular Biology, Volumes I, II, and III, 1997 (F. M. Ausubel ed.); Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; DNA Cloning: A Practical Approach, Volumes I and II, 1985 (D. N. Glover ed.); Oligonucleotide Synthesis, 1984 (M. L. Gait ed.); Nucleic Acid Hybridization, 1985, (Hames and Higgins); Transcription and Translation, 1984 (Hames and Higgins eds.); Animal Cell Culture, 1986 (R. I. Freshney ed.); Immobilized Cells and Enzymes, 1986 (IRL Press); Perbal, 1984, A Practical Guide to Molecular Cloning; the series, Methods in Enzymology (Academic Press, Inc.); Gene Transfer Vectors for Mammalian Cells, 1987 (J. H. Miller and M. P. Calos eds., Cold Spring Harbor Laboratory); and Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively).

As used herein, "differentially expressed gene" refers to (a) a gene containing at least one of the DNA sequences disclosed herein (e.g., as shown in Figure 1 and SEQ ID NO:1); (b) any DNA sequence that encodes the amino acid sequence encoded by the DNA sequences disclosed herein (e.g., as shown in Figure 2 and SEQ ID NO:2); or (c) any DNA sequence that is substantially similar to the coding sequences disclosed herein.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence, e.g. where only changes in amino acids not affecting the polypeptide function occur. Desirably the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide sequence. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence desirably is at least 80%, more desirably at least 85%, preferably at least 90%, more preferably at least 95%, still more preferably at least 99%.

Sequence comparisons are carried out using a Smith-Waterman sequence alignment algorithm (see e.g. Waterman, M.S. Introduction to Computational Biology: Maps, sequences and genomes. Chapman & Hall. London: 1995. ISBN 0-412-99391-0, or at http://www-hto.usc.edu/software/seqaln/index.html). The localS program, version 1.16, is used with following parameters: match: 1, mismatch penalty: 0.33, open-gap penalty: 2, extended-gap penalty: 2. A nucleotide sequence "substantially similar" to reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C, yet still encodes a functionally equivalent gene product.

The differentially expressed genes disclosed herein are expressed in arthritic tissues (e.g., cartilage in a human afflicted with arthritis) in elevated amounts relative to, i.e., to a greater extent than in the corresponding tissues of humans who do not suffer from arthritis. Messenger RNA transcribed from the differentially expressed genes, and protein translated from such mRNA, is present in arthritic tissues and/or synovial fluid associated with such tissues in an amount at least about twice, preferably at least about five times, more preferably at least amount ten times, most preferably at least about 100 times the levels of mRNA and protein found in corresponding tissues found in humans who do not suffer from arthritis. Such elevated transcription of adlican mRNA is referred to herein as "elevated transcription."

A "host cell," as used herein, refers to a prokaryotic or eukaryotic cell that contains heterologous DNA that has been introduced into the cell by any means, e.g., electroporation, calcium phosphate precipitation, microinjection, transformation, viral infection, and the like.

"Heterologous" as used herein means "of different natural origin" or represent a non-natural state. For example, if a host cell is transformed with a DNA or gene derived from another organism, particularly from another species, that gene is heterologous with respect to

that host cell and also with respect to descendants of the host cell which carry that gene. Similarly, heterologous refers to a nucleotide sequence derived from and inserted into the same natural, original cell type, but which is present in a non-natural state, e.g. a different copy number, or under the control of different regulatory elements.

5   A vector molecule is a DNA molecule into which heterologous DNA may be inserted which can then replicate in an appropriate host cell. Vectors must have one or more origin of replication, and one or more site into which the recombinant DNA can be inserted. Vectors often have convenient means by which cells with vectors can be selected from those without, e.g., they encode drug resistance genes. Common vectors include plasmids, viral genomes, and (primarily

10   in yeast) "artificial chromosomes."

"Plasmids" generally are designated herein by a lower case p preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Starting plasmids disclosed herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available

15   plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art. Moreover, those of skill readily may construct any number of other plasmids suitable for use in the invention. The properties, construction and use of such plasmids, as well as other vectors, in the present invention will be

20   readily apparent to those of skill from the present disclosure.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the

25   natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

As used herein, the term "transcriptional control sequence" refers to DNA sequences, such as initiator sequences, enhancer sequences, and promoter sequences, which induce, repress,

or otherwise control the transcription of protein encoding nucleic acid sequences to which they are operably linked.

As used herein, "human adlican transcriptional control sequences" are any of those transcriptional control sequences normally found associated with the human adlican gene in as it is found in the human X chromosome.

As used herein, "non-human transcriptional control sequence" is any transcriptional control sequence not found in the human genome.

The term "polypeptide" is used interchangeably herein with the terms "polypeptides" and "protein(s)".

The invention also includes nucleic acid molecules, preferably DNA molecules, that hybridize to, and are therefore the complements of, the DNA sequences (a) through (c), in the preceding paragraph. Such hybridization conditions may be highly stringent or less highly stringent, as described above. In instances wherein the nucleic acid molecules are deoxyoligonucleotides ("oligos"), highly stringent conditions may refer, e.g., to washing in 6.times.SSC/0.05% sodium pyrophosphate at 37 °C. (for 14-base oligos), 48 °C. (for 17-base oligos), 55 °C. (for 20-base oligos), and 60 °C. (for 23-base oligos). These nucleic acid molecules may act as target gene antisense molecules, useful, for example, in target gene regulation and/or as antisense primers in amplification reactions of target gene nucleic acid sequences. Further, such sequences may be used as part of ribozyme and/or triple helix sequences, also useful for target gene regulation. Still further, such molecules may be used as components of diagnostic methods whereby the presence of a cardiovascular disease-causing allele, may be detected.

The invention also encompasses (a) DNA vectors that contain any of the foregoing coding sequences and/or their complements (i.e., antisense); (b) DNA expression vectors that contain any of the foregoing coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences; and (c) genetically engineered host cells that contain any of the foregoing coding sequences operatively associated with a regulatory element that directs the expression of the coding sequences in the host cell. As used herein, regulatory elements include but are not limited to inducible and non-inducible promoters, enhancers, operators and other elements known to those skilled in the art that drive and regulate expression.

The invention includes fragments of any of the DNA sequences disclosed herein. Fragments of the full length adlican gene may be used as a hybridization probe for a cDNA library to isolate the full length gene and to isolate other genes which have a high sequence similarity to the adlican gene or similar biological activity. Probes of this type preferably have at

5   least about 30 bases and may contain, for example, from about 30 to about 50 bases, about 50 to about 100 bases, about 100 to about 200 bases, or more than 200 bases. The probe may also be used to identify a cDNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete ADLICAN gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the

10   ADLICAN gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

In addition to the gene sequences described above, homologs of such sequences, as may,

15   for example be present in other species, may be identified and may be readily isolated, without undue experimentation, by molecular biological techniques well known in the art. Further, there may exist genes at other genetic loci within the genome that encode proteins which have extensive homology to one or more domains of such gene products. These genes may also be identified via similar techniques.

20   For example, the isolated differentially expressed gene sequence may be labeled and used to screen a cDNA library constructed from mRNA obtained from the organism of interest. Hybridization conditions will be of a lower stringency when the cDNA library was derived from an organism different from the type of organism from which the labeled sequence was derived. Alternatively, the labeled fragment may be used to screen a genomic library derived from the

25   organism of interest, again, using appropriately stringent conditions. Such low stringency conditions will be well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Springs Harbor Press, N.Y.; and Ausubel et al., 1989, Current

30   Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

Further, a previously unknown differentially expressed gene-type sequence may be isolated by performing PCR using two degenerate oligonucleotide primer pools designed on the basis of amino acid sequences within the gene of interest. The template for the reaction may be cDNA obtained by reverse transcription of mRNA prepared from human or non-human cell lines or tissue known or suspected to express a differentially expressed gene allele.

The PCR product may be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a differentially expressed gene-like nucleic acid sequence. The PCR fragment may then be used to isolate a full length cDNA clone by a variety of methods. For example, the amplified fragment may be labeled and used to screen a bacteriophage cDNA library. Alternatively, the labeled fragment may be used to screen a genomic library.

PCR technology may also be utilized to isolate full length cDNA sequences. For example, RNA may be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction may be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis. The resulting RNA/DNA hybrid may then be "tailed" with guanines using a standard terminal transferase reaction, the hybrid may be digested with RNAase H, and second strand synthesis may then be primed with a poly-C primer. Thus, cDNA sequences upstream of the amplified fragment may easily be isolated. For a review of cloning strategies which may be used, see e.g., Sambrook et al., 1989, supra.

In cases where the differentially expressed gene identified is the normal, or wild type, gene, this gene may be used to isolate mutant alleles of the gene. Such an isolation is preferable in processes and disorders which are known or suspected to have a genetic basis. Mutant alleles may be isolated from individuals either known or suspected to have a genotype which contributes to cardiovascular disease symptoms. Mutant alleles and mutant allele products may then be utilized in the diagnostic assay systems described below.

A cDNA of the mutant gene may be isolated, for example, by using PCR, a technique which is well known to those of skill in the art. In this case, the first cDNA strand may be synthesized by hybridizing an oligo-dT oligonucleotide to mRNA isolated from tissue known or suspected to be expressed in an individual putatively carrying the mutant allele, and by extending the new strand with reverse transcriptase. The second strand of the cDNA is then synthesized

using an oligonucleotide that hybridizes specifically to the 5' end of the normal gene. Using these two primers, the product is then amplified via PCR, cloned into a suitable vector, and subjected to DNA sequence analysis through methods well known to those of skill in the art. By comparing the DNA sequence of the mutant gene to that of the normal gene, the mutation(s) responsible for the loss or alteration of function of the mutant gene product can be ascertained.

Alternatively, a genomic or cDNA library can be constructed and screened using DNA or RNA, respectively, from a tissue known to or suspected of expressing the gene of interest in an individual suspected of or known to carry the mutant allele. The normal gene or any suitable fragment thereof may then be labeled and used as a probed to identify the corresponding mutant allele in the library. The clone containing this gene may then be purified through methods routinely practiced in the art, and subjected to sequence analysis as described above.

Additionally, an expression library can be constructed utilizing DNA isolated from or cDNA synthesized from a tissue known to or suspected of expressing the gene of interest in an individual suspected of or known to carry the mutant allele. In this manner, gene products made by the putatively mutant tissue may be expressed and screened using standard antibody screening techniques in conjunction with antibodies raised against the normal gene product, as described, below. (For screening techniques, see, for example, Harlow, E. and Lane, eds., 1988, "Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor.) In cases where the mutation results in an expressed gene product with altered function (e.g., as a result of a missense mutation), a polyclonal set of antibodies are likely to cross-react with the mutant gene product. Library clones detected via their reaction with such labeled antibodies can be purified and subjected to sequence analysis as described above.

Differentially expressed gene products include those proteins encoded by nucleotide sequence set forth in SEQ ID NO: 1, in particular, a polypeptide that is or includes the amino acid sequence set out in SEQ ID NO: 2, or fragments thereof.

In addition, differentially expressed gene products may include proteins that represent functionally equivalent gene products. Such an equivalent differentially expressed gene product may contain deletions, additions or substitutions of amino acid residues within the amino acid sequence encoded by the differentially expressed gene sequences described, above, but which result in a silent change, thus producing a functionally equivalent differentially expressed on

pathway gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved.

For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Functionally equivalent", as utilized herein, refers to a protein capable of exhibiting a substantially similar in vivo activity as the endogenous differentially expressed gene products encoded by the differentially expressed gene sequences described above. Alternatively, when utilized as part of assays such as those described, below, "functionally equivalent" may refer to peptides capable of interacting with other cellular or extracellular molecules in a manner substantially similar to the way in which the corresponding portion of the endogenous differentially expressed gene product would.

The differentially expressed gene products may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing the differentially expressed gene polypeptides and peptides of the invention by expressing nucleic acid encoding differentially expressed gene sequences are described herein. Methods which are well known to those skilled in the art can be used to construct expression vectors containing differentially expressed gene protein coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques and in vivo recombination/genetic recombination. See, for example, the techniques described in Sambrook et al., 1989, supra, and Ausubel et al., 1989, supra. Alternatively, RNA capable of encoding differentially expressed gene protein sequences may be chemically synthesized using, for example, synthesizers. See, for example, the techniques described in "Oligonucleotide Synthesis", 1984, Gait, M. J. ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety.

A variety of host-expression vector systems may be utilized to express the differentially expressed gene coding sequences of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified,

but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, exhibit the differentially expressed gene protein of the invention in situ. These include but are not limited to microorganisms such as bacteria (e.g., E. coli, B. subtilis) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA

5 expression vectors containing differentially expressed gene protein coding sequences; yeast (e.g. Saccharomyces, Pichia) transformed with recombinant yeast expression vectors containing the differentially expressed gene protein coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the differentially expressed gene protein coding sequences; plant cell systems infected with recombinant virus expression

10 vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing differentially expressed gene protein coding sequences; or mammalian cell systems (e.g. COS, CHO, BHK, 293, 3T3) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothioneine promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter).

Expression of adlican by a cell from an adlican gene that is native to a the cell can also be performed. Methods for such expression are detailed in, e.g., U.S. Patents 5,641,670; 5,733,761; 5,968,502; and 5,994,127, all of which are expressly incorporated by reference herein in their entirety.

20 In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the differentially expressed gene protein being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of antibodies or to screen peptide libraries, for example, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include,

25 but are not limited, to the *E. coli* expression vector pUR278 (Ruther et al., 1983, EMBO J. 2:1791), in which the differentially expressed gene protein coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye & Inouye, 1985, Nucleic Acids Res. 13:3101-3109; Van Heeke & Schuster, 1989, J. Biol. Chem. 264:5503-5509); and the like. PGEX vectors may also be used to

30 express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general,

such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene protein can be released from the GST moiety.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("cat") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available. Two such vectors are pKK232-8 and pCM7. Thus, promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the E. coli lacI and lacZ promoters, the T3 and T7 promoters, the T5 tac promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is one of several insect systems that can be used as a vector to express foreign genes. The virus grows in Spodoptera frugiperda cells. The differentially expressed gene coding sequence may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of differentially expressed gene coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used

15

to infect Spodoptera frugiperda cells in which the inserted gene is expressed. (E.g., see Smith et al., 1983, J. Virol. 46: 584; Smith, U.S. Pat. No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the differentially expressed gene coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing differentially expressed gene protein in infected hosts. (E.g., See Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted differentially expressed gene coding sequences. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire differentially expressed gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the differentially expressed gene coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner et al., 1987, Methods in Enzymol. 153:516-544).

Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for expression vector construction, introduction of the vector into the host and expression in the host per se are routine skills in the art.

Generally, recombinant expression vectors will include origins of replication, a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector.

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, etc.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the differentially expressed gene protein may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the differentially expressed gene protein. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the differentially expressed gene protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, et al., 1977, Cell 11:223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy, et al., 1980, Cell 22:817) genes can be employed in tk⁻, hgprt⁻ or aprt⁻ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, et al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare, et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers

resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, et al., 1981, J. Mol. Biol. 150:1); and hygro, which confers resistance to hygromycin (Santerre, et al., 1984, Gene 30:147) genes.

An alternative fusion protein system allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (Janknecht, et al., 1991, Proc. Natl. Acad. Sci. USA 88: 8972-8976). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+} nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

When used as a component in assay systems such as those described below, the differentially expressed gene protein may be labeled, either directly or indirectly, to facilitate detection of a complex formed between the differentially expressed gene protein and a test substance. Any of a variety of suitable labeling systems may be used including but not limited to radioisotopes such as ^{125}I; enzyme labeling systems that generate a detectable calorimetric signal or light when exposed to substrate; and fluorescent labels.

Where recombinant DNA technology is used to produce the differentially expressed gene protein for such assay systems, it may be advantageous to engineer fusion proteins that can facilitate labeling, immobilization and/or detection.

Indirect labeling involves the use of a protein, such as a labeled antibody, which specifically binds to either a differentially expressed gene product. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by an Fab expression library.

Described herein are methods for the production of antibodies capable of specifically recognizing one or more differentially expressed gene epitopes. Such antibodies may include, but are not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Such antibodies may be used, for example, in the detection of a fingerprint, target,

gene in a biological sample, or, alternatively, as a method for the inhibition of abnormal target gene activity. Thus, such antibodies may be utilized as part of cardiovascular disease treatment methods, and/or may be used as part of diagnostic techniques whereby patients may be tested for abnormal levels of adlican, or for the presence of abnormal forms of adlican.

For the production of antibodies to a differentially expressed gene, various host animals may be immunized by injection with a differentially expressed gene protein, or a portion thereof. Such host animals may include but are not limited to rabbits, mice, and rats, to name but a few. Various adjuvants may be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen, such as target gene product, or an antigenic functional derivative thereof. For the production of polyclonal antibodies, host animals such as those described above, may be immunized by injection with differentially expressed gene product supplemented with adjuvants as also described above.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to the hybridoma technique of Kohler and Milstein, (1975, Nature 256:495-497; and U.S. Pat. No. 4,376,110), the human B-cell hybridoma technique (Kosbor et al., 1983, Immunology Today 4:72; Cole et al., 1983, Proc. Natl. Acad. Sci. USA 80:2026-2030), and the EBV-hybridoma technique (Cole et al., 1985, Monoclonal Antibodies And Cancer Therapy, Alan R. Liss, Inc., pp. 77-96). Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo makes this the presently preferred method of production.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., 1984, Proc. Natl. Acad. Sci., 81:6851-6855; Neuberger et al., 1984, Nature, 312:604-608;

Takeda et al., 1985, Nature, 314:452-454) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived

5   from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, 1988, Science 242:423-426; Huston et al., 1988, Proc. Natl. Acad. Sci. USA 85:5879-5883; and Ward et al., 1989, Nature 334:544-546) can be adapted to produce differentially expressed gene-single chain antibodies. Single chain antibodies are formed by

10   linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments

15   which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed (Huse et al., 1989, Science, 246:1275-1281) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Particularly preferred, for ease of detection, is the sandwich assay, of which a number of variations exist, all of which are intended to be encompassed by the present invention.

20   For example, in a typical forward assay, unlabeled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen binary complex. At this point, a second antibody, labeled with a reporter molecule capable of inducing a detectable signal, is then added and incubated, allowing time sufficient for

25   the formation of a ternary complex of antibody-antigen-labeled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal, or may be quantitated by comparing with a control sample containing known amounts of antigen. Variations on the forward assay include the simultaneous assay, in which both sample and antibody are added simultaneously to the bound antibody, or a reverse assay in which the labeled

30   antibody and sample to be tested are first combined, incubated and added to the unlabeled

surface bound antibody. These techniques are well known to those skilled in the art, and the possibility of minor variations will be readily apparent. As used herein, "sandwich assay" is intended to encompass all variations on the basic two-site technique. For the immunoassays of the present invention, the only limiting factor is that the labeled antibody be an adlican-specific

5   antibody.

The most commonly used reporter molecules in this type of assay are either enzymes, fluorophore- or radionuclide-containing molecules. In the case of an enzyme immunoassay an enzyme is conjugated to the second antibody, usually by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different ligation techniques exist,

10   which ae well-known to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, among others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable color change. For example, p-nitrophenyl phosphate is suitable for use with alkaline phosphatase conjugates; for peroxidase

15   conjugates, 1,2-phenylenediamine or toluidine are commonly used. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. A solution containing the appropriate substrate is then added to the tertiary complex of antibody-adlican-labeled antibody. The substrate reacts with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually

20   spectrophotometrically, to give an evaluation of the amount of adlican which is present in the serum sample.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labeled antibody absorbs the

25   light energy, inducing a state of excitability in the molecule, followed by emission of the light at a characteristic longer wavelength. The emission appears as a characteristic color visually detectable with a light microscope. Immunofluorescence and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotopes, chemiluminescent or bioluminescent molecules may

also be employed. It will be readily apparent to the skilled artisan how to vary the procedure to suit the required use.

The present invention also includes a kit for performing the assay aspect of the invention. Such a kit includes vials or vessels for incubating a body tissue or fluid sample, and the components necessary for quantifying human adlican polynucleotides, for example, via RT-PCR. A kit for quantifying human adlican polypeptide may contain anti-adlican antibodies, for example, the antibodies may be prepared via the procedure set forth in Example 5.

The following Examples illustrate the present invention, without in any way limiting the scope thereof.

EXAMPLES

Example 1: Cloning of Adlican

Differential display RT-PCR is used to identify gene products with altered expression in diseased cartilage from humans suffering from OA compared to cartilage from non-arthritic (NA) human controls. As described by Hu *et al.*, above, first strand cDNA is synthesized from 0.2 μg of total RNA with each of the 3 anchored oligo-dT primers from GenHunter Corporation, Nashville, TN. The reaction (20 μl) was carried out at 37 °C for 60 min. For PCR amplification, 1 μl of the cDNA serves as template in a 10 μl reaction mix containing 10 mM Tris-HCl (pH 8.4), 1.5 mM $MgCl_2$, 50 mM KCl, 0.001% gelatin, 2 μM dNTPs, 0.2 μM of 5' arbitrary primer (AP-3 from the RNAimage kit obtained from GenHunter Corporation, Nashville, TN), 2 μM of the same anchored primer used in the cDNA synthesis, 5 μCi of α-[^{33}P]dATP (2,000 Ci/mmole, Dupont-New England Nuclear, Boston, MA) and 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer, Norwalk, CT). Samples are subjected to 40 cycles of amplification under the following conditions: denaturing at 94 °C, 30 sec, annealing at 40 °C, 2 min, extension at 72 °C, 30 sec, and a final extension at 72 °C, 5 min. The resulting PCR products are resolved on a denaturing polyacrylamide gel and visualized by autoradiography of the dried gel. PCR products of interest are excised from the gel, and the DNA is eluted and re-amplified by PCR using the same primers and conditions described above, excluding the radio-labeled nucleotide. PCR products are analyzed on a 1.5% agarose gel and are ligated into the cloning vector PCR II 2.1

(TA cloning Kit, Invitrogen, Carlsbad, CA). Clones of the PCR-generated fragments are obtained by transformation of *E. coli* strain DH5α (Gibco/BRL, Gaithersberg, MD). DNA sequences are determined for at least 3 independent clones of each fragment using Dye Terminator Cycle Sequencing on an ABI PRISM 377 DNA sequencing system (Perkin-Elmer, Norwalk, CT).

As described by Hu *et al*, above, first strand cDNA is synthesized from total RNA isolated from OA and non-arthritic cartilage. 200 ng of total RNA and 10 pmoles of primer $T_{30}VN$ (where V=A,C,G and N=A,C,G,T) are mixed in a 6 μl volume, heated to 72°C for 3 min and quenched on ice for 3 minutes Buffer and MMLV reverse transcriptase are added to final concentrations of 50 mM Tris-HCl (pH 8.3), 6 mM $MgCl_2$, 75 mM KCl, 1 mM dNTPs, and 10 units MMLV reverse transcriptase in 10 μl. This mixture is incubated at 42 °C for 1.5 hr, 94 °C for 5 min, and quenched on ice. Serial dilutions of cDNA from different individuals are used for PCR amplification with a primer set for actin (forward: GGAGTCCTGTGGCATCCACGAAACTAC (SEQ ID NO:3) and reverse: CACATCTGCTGGAAGGTGGACAGCG (SEQ ID NO:4) and for 3G12 (forward: CATGGGCACAAACTGACTCATGGCTG (SEQ ID NO:5) and reverse: GAGAGGAGAGGAAGGAGAAAACTGCATC (SEQ ID NO:6)) under the following conditions: 25 μl reaction volume with 10 mM Tris-HCl (pH 8.3) , 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 20 μM of dNTPs, 1.25 units AmpliTaq Gold Polymerase (Perkin Elmer, Norwalk, CT); 94 °C, 8.5 minutes; 32 cycles of 94 °C, 30 seconds; 63 °C, 30 seconds; 72 °C, 2 minutes, and a final incubation at 72 °C, 7 minutes. Nine μl of the reaction mix is run on a 10% polyacrylamide gel, stained with SYBR™ Green I (Molecular Probes, Eugene, OR), and quantified using fluorescence imaging and ImageQuant software (Molecular Dynamics, Sunnyvale, CA). The resulting PCR fragment for 3G12 is 155 bp. Concentrations are chosen to ensure that the reaction is within the log phase of amplification. Levels of expression are normalized to the levels of actin for each sample. A negative control reaction with no template is carried out for each primer set to verify the absence of contamination.

From this screening approach, several differentially expressed transcripts are obtained, including a 251 bp clone designated 3G12 (detected with 5' primer AP-3 and 3' primer oligo dT with G, 12th clone isolated). The sequence of 3G12 is determined to be a novel sequence identified in the screen. Using PCR primers designed based on the 3G12 sequence, the

corresponding transcript is found to be expressed over 100 times greater in OA cartilage than in NA cartilage by semi-quantitative RT-PCR (Figure 3). Because of the high degree of differential expression, 3G12 is pursued as a diagnostic marker gene for OA.

The cloned 251 bp PCR-generated 3G12 fragment is used as a probe to identify a λgt11 phage clone designated "106A" with a 1.4 kb insert in a human placenta cDNA library. From a human placental 5'-STRETCH PLUS cDNA Library (Oligo dT + random primed, Clontech, Palo Alto, CA), 10^7 pfu were divided into 40 sub-libraries and screened by PCR using 3G12-specific primers. Positive fractions were further divided into 21 second-round sub-libraries which were also screeened by PCR. Phage from the positive second-round sub-libraries were plated on *E. coli* strain Y1090 and screened by plaque hybridisation using a PCR-generated probe (251 bp) labeled with $[^{32}P]$ dATP. Filters were hybridized at 38 °C for 2.5 hours and washed at 48 °C three times in 1x SSC containing 0.1% SDS. Positive plaques were identified by autoradiography, isolated and amplified, and re-screened until positive single phage clones were confirmed.

A BLASTX analysis of the sequence of 106A identifies a partial open reading frame (ORF) with coding sequence related to the L1 neural cell adhesion molecule (Coutelle, O., Nyakatura, G., Taudien, S., Elgar, G., Brenner, S., Platzer, M., Drescher, B., Jouet, M., Kenwrick, S., and Rosenthal, A. (1998) *Gene* **208**: 7-15), neurofascin (Hassel, B., Rathjen, F.G., and Volkmer, H. (1997) *J Biol Chem* **272**: 28742-28749), and to perlecan, a large heparan sulfate containing proteoglycan (Murdoch, A.D., Dodge, G.R., Cohen, I., Tuan, R.S., Iozzo, R.V. (1992) *J Biol Chem* **267**: 8544-8557). Although it is evident that 106A represents a partial transcript, no cDNA clones are yet identified with additional upstream sequences. A repeated search of the Genbank and the EST sequence databases is conducted, and a small number of EST sequences corresponding to 3G12 are found. However, none of these sequences include the 5'-segment of cDNA clone 106A.

Clone 106A is used to identify a human genomic clone (Genome Systems, St. Louis, MO, P1 clone # 14414). Sub-clones of the P1 isolate are obtained, are characterized by DNA sequence analysis, and are found to contain 2 complete and 1 incomplete exons. The 5'-most exon region is then used as a probe to identify another cDNA clone from the placental library,

designated N3. The sequence of N3 extends the ORF, revealing additional homologies to NCAM and perlecan; however, the sequence lacks any evident translational initiation site.

An approach known as 5' RACE (rapid amplification of cDNA ends) is then used to obtain extended clones. Separate first strand cDNA syntheses are carried out with 500 ng of total RNA isolated and pooled from 4 OA cartilage samples. Aliquots are combined for the synthesis of double stranded cDNA with anchors ligated at both ends according to the Marathon™ cDNA system (Clontech, Palo Alto, CA). The antisense primer is derived from the sequence of cDNA clone N3. Utilization of the 5' RACE method yields an additional ~500 bp of coding sequence, although still with no evident translational start site. The 5'-most sequence from the RACE clones is used to isolate a second genomic clone, and its sub-clone E3. A search of Genbank reveals that the sequence of E3 is 100% identical to a 3 kb segment of a sequence deposited into Genbank (Accession number AC004616,), human genomic clone BAC536K7 (149,752 bp) containing a segment of the X chromosome.

The sequence of BAC536K7 is extensively analysed for predicted exons encoding the transcript corresponding to clone 3G12. Various exon prediction programs (e.g. GENESCAN, GRAIL) are used to find the rest of the gene encoding 3G12. PCR experiments are required to test for the presence of putative exons using the original placental cDNA library as template. The 5' exon (X1, exons are referred to 5' to 3' in their order in the gene, X1, X2, etc.) predicted by GENESCAN to contain the translation start site is found to be incorrect. Only by visual inspection using ORF FINDER of small potential coding exons, which contain apparent secretory signal sequences connected to sequences homologous to proteins in common with a down stream exon is the correct 5' exon identified. Exons X1, X2, X3, and X4 are verified by PCR using human placental cDNA as template and primers which bridged pairs of exons. The complete sequence of the 3G12 transcript is derived from 6 exons (X5 and X6 represented by clones N3 and 106A). The entire DNA sequence and encoded protein sequence are shown in Figures 1 and 2, respectively. The large adlican protein (2,828 amino acids) contains five leucine-rich repeats in the amino terminal region downstream of the signal sequence, distinct regions of homology to neural cell adhesion molecule and perlecan containing immunoglobulin C2 domains, and a mucin-like domain. Based on its predicted functionalities, the protein is

termed "adlican" – for *AD*hesion protein with *L*eucine-rich repeats and *I*mmunoglobulin domains related to perle*CAN*.

5

Example 2: Assembly of Expression Construct for Adlican

In order to produce adlican protein and examine its biochemical properties, multiple sub-cloning steps are required to assemble the various cDNA and genomic clone fragments, comprising the entire coding region, into an expression vector. A first construct is made as

10 follows. A λgt11 clone which contains 347 bp of 5' UTR and 1-715 bp of coding region is used as a template for PCR. The PCR primer sequences used were:

Upstream (contains a NotI site and the initiation codon ATG)

5'- TTGCGGCCGCGCCACCATGCCCAAGCGCGCGCACTGG-3' (SEQ ID NO:7);

and

15 Downstream

5'-TCAATACTCCTGCTCCTGTTCTGTCTCA-3' (SEQ ID NO:8).

During the PCR reaction a NotI site is added at the 5'-end to enable subcloning into an expression vector and the 3'-end contained an EcoRI site which is present in the original adlican sequence. The 715bp NotI-EcoRI fragment is subcloned into the pcDNA 3 vector (Invitrogen,

20 Carlsbad, CA) and the nucleotide sequence is verified using an Applied Biosystems 373 automated DNA sequencer and dideoxy terminator chemistry. This fragment also contains a signal sequence of 19 amino acids.

The following cDNA clones are used to generate a second construct. Clone E3 which contains coding regions from 1513-5482 bp is cut with HindIII and EcoRI and a 3969 bp

25 fragment which corresponds to base pairs 1513-5162 of adlican is subcloned into the pcDNA3 vector. A HindIII-HindIII fragment which contains sequences corresponding to base pairs 734-1512bp (778bp) of adlican is generated by PCR using the D5 clone as a template, subcloned and sequenced. The vector that contains the 3969 base pair EcoRI- HindIII fragment is cut with HindIII and the 778 bp HindIII-HindIII fragment is ligated into it. The clone containing the

30 HindIII fragment in the correct orientation contains cDNA sequence corresponding to base pairs

734-5162 of the original sequence. A 19 base pair EcoRI – HindIII adaptor which contains sequence corresponding to base pairs 716-733 is synthesized and ligated to construct II which is partially digested with HindIII to open up the 5' HindIII site located at bp734. This completes construction of the second construct, which contains an insert of 4446 bp (base pairs 716-5162) of the original sequence.

The third construct is generated from E3 and two other cDNA clones N3 and 106A. E3 is cut with EcoRI-HindIII to generate a 320 bp fragment corresponding to base pairs 5163-5482 of adlican. N3 is cut with HindIII-ApaLI to generate a 297 bp fragment corresponding to base pairs 5483-5780 of adlican. Clone 106A is cut with ApaLI- XhoI to generate a 2707 bp fragment which corresponds to base pairs 5780-8487 of adlican. This fragment also contains a stop codon at base pair 8484. The EcoRI- HindIII, HindIII- ApaLI and ApaLI-XhoI fragments are put together by a trimolecular ligation and are subcloned into EcoRI- XhoI cut pcDNA3 to generate a 3325 bp fragment, i.e., the third construct, corresponding to base pairs 5162-8487 of adlican.

In order to generate the full length clone, the third construct is cut with EcoRI and NotI and the NotI-EcoRI insert of the first construct from above is ligated into it to generate a fourth construct of 4040 bp. The second construct from above is cut with EcoRI and the 4446 bp insert is gel purified.

The fourth construct is digested with EcoRI and the 4446 bp EcoRI insert of the second construct is ligated into it to generate the final full length fifth construct. Clones which have the EcoRI fragment of the second construct in the right orientation represent the full length clone for adlican. The stop codon is removed from the full length insert and further subcloned in frame into a hexa-histidine tag containing vector (Invitrogen, Carlsbad, CA) for *in vitro* and *in vivo* expression.

Example 3: Expression of the full-length construct in a TNT reticulocyte lysate coupled Transcription/Translation system

The plasmid DNA containing the full length insert is translated in a TNT® coupled reticulocyte lysate system (Promega, Madison, WI) as described by the manufacturer. The reaction mixture containing TNT® rabbit reticulocyte lysate, TNT® reaction buffer, T7 RNA polymerase, amino acid mixture minus methionine, [^{35}S]- methionine (> 1000Ci/mmole) Rnasin-

ribonuclease inhibitor, 0.5µg of pcDNA 3.1 (Invitrogen, Carlsbad, CA), DNA containing the full length adlican insert and transcend biotin-lysyl-tRNA is incubated at 30 °C for 60-90 minutes. At the end of the incubation an aliquot of the reaction is denatured in SDS containing sample buffer and electrophoresed on a 4-20% gradient polyacrylamide gel. The gel is dried after

5   soaking for 5 minutes in 7% acetic acid, 7% methanol, 1% glycerol containing buffer and translated products are visualized by autoradiography.

Example 4: Transient expression of the full-length construct in HEK 293 cells

HEK 293 cells are seeded (200,000/well) into 6 well plates in DMEM medium containing

10   10% FBS and are incubated at 37 °C for 48-72 hours until the cells are about 70% confluent. The cells are transfected with the full length adlican construct in the pcDNA 3.1 vector using lipofectamine/Plus reagent (Gibco/BRL, Gaithersberg, MD). The DNA is precomplexed with the Plus reagent for 15 minutes and then is incubated with diluted lipofectamine reagent for a further 15 minutes. The cells are washed and transferred to serum-free medium and the

15   DNA/lipofectamine mixture is added and incubated for 3 hours at 37 °C. The transfection medium is removed and is replaced with complete medium (DMEM containing 10% FBS). The cells are incubated overnight. The cells are shifted into DMEM containing 0.2% FBS the next day and are incubated for a further 48 hours. Aliquots of the supernatant are collected 48 hours and 72 hours post-transfection, are denatured and are electrophoresed on a 4-20% gradient

20   polyacrylamide gel. The proteins in the gel were electroblotted on to a PVDF membrane (BioRad, Hercules, CA) and probed with anti hexa-his antibody (Qiagen, Valencia, CA). The bound antibody was visualized using the ECL reagent system (Amersham, Piscataway, NJ).

Example 5: Generation of Antibodies that Recognize and Bind to Adlican

25   An anti-peptide antibody is generated against a sequence near the carboxy-terminus of adlican based on the predicted coding sequence of the first cDNA clone, 106A. A 15 residue peptide (CMAKNILGSDSKTTY), corresponding to the sequence of the adlican protein near the carboxyl terminus, is designed based on surface probability as determined using the program PROTEAN, a component of the LASERGENE suite of programs (DNASTAR Inc., Madison,

30   WI). The peptide is synthesized, purified, and used to immunize two rabbits following a standard

protocol (Genosys Biotechnologies, The Woodlands, TX). Antiserum from one of the immunized animals which shows the highest titer against the peptide is used for immunoblot analysis. Samples (4 μl) of human synovial fluid are subjected to SDS polyacrylamide gel electrophoresis, are transferred to PVDF membranes (BioRad, Hercules, CA), are analyzed for immunoreactivity using a 1:500 dilution of the antiserum, using the ECL detection system (chemiluminescent) (Amersham, Piscataway, NJ).

The resulting antiserum is used for detecting adlican protein in human synovial fluid by immunoblot analysis. The data in Fig. 6 include an example of one of the immunoblots where adlican protein fragment is detected in human synovial fluid and a tabulation of results derived from several blots where synovial fluids from individual patients with varied diagnoses are tested. The diagnoses are osteoarthritis, rheumatoid arthritis, gout, and joint trauma. 100% of the patients with severe OA show a positive adlican protein immunoreactivity.